

LITERATURE SURVEY

GUI For Analysis

Team 6

Dhanushkumar S G 20BCS026

Dhinakaran S B 20BCS029

Gokul M 20BCS032

Karthikeyan K 20BCS050

Karthikeyan R 20BCS051

Problem Statement

Data analytics is critical since it allows firms to improve their performance. Cleaning the data involves the removal of duplicates, the removal or replacement of missing items, the correction of misfielded values, the maintenance of uniform formatting, and a variety of additional duties. Having clean data will eventually boost overall productivity and provide for the greatest quality information in your decision-making, which takes time. As a result, an automatic data cleansing tool is necessary.

<div>Literature Survey</div>	<div>Problem Statement:</div>	<div>Many users found difficulties in analyzing the data . The visualizing and analyzing the data might be difficult for general purpose. They are in need for a powerful research data analysis tool.</div>
<div>Research Data Analysis with Power BI</div> <div>Vijay Krishnan S Bharanidharan G Krishnamoorthy</div>	<div>Proposed Solution:</div>	<div>Power BI is a Data Visualization and Business Intelligence tool that converts data from different data sources to interactive dashboards and BI reports.</div>
	<div>Output:</div>	<div>Power BI, from Microsoft, is a suite of business analytics tools that is used to analyze data and share insights in the form of reports and dashboards. User data in various forms – spreadsheets, text files, databases, etc. form the input for Power BI. Datasets are formed by transforming the data provided by the users. Data transformations are decided by the users. This step is used to remove errors and redundant data, correct formatting, and prepare data for further analysis by organizing them into suitable normalized forms, and so on. Based on the report and dashboard being developed, filtering the data to only include the relevant bits enables one to focus on only the data that matters.</div>
	<div>Merits:</div>	<div>Power BI makes it easy to bring data to life. Users can connect their data to Power BI and choose from a variety of visualizations (column charts, pie charts, bubble and heat maps, scatter plots, etc.) to tell stories about their data in order to gain and share insights.</div>
	<div>Demerits:</div>	<div>Complex to Understand and Master Usually, Power BI is the easiest to use BI tool if you are using it simply to import data and create reports. But Power BI is an entire suite having a lot of other interrelated tools. When the purpose of your use is more than just creating reports in Power BI Desktop, you need to learn and master several other tools like <i>Gateways</i>, <i>Power BI Report Server</i>, <i>Power BI</i></div>

<div>Literature Survey</div> <div>TITLE AND AUTHOR</div>	<div>Problem Statement:</div>	<div>Many dataset comes with incomplete, blank, inaccurate data. To conduct data analysis, it is required to remove the numerous datasets that contain corrupted data.</div>
<div>Automated Data Cleaning</div> <div>Jowin Jestine</div> <div>Anushka Kamath</div> <div>Maroof Khatib</div> <div>Priya Nawal</div>	<div>Proposed Solution:</div>	<div>ADC (Automated Data Cleaning) tries to give the accuracy and precision of the dataset provided by the user. Efficiency of the cleaned dataset is checked against the accuracy, precision and recall of the model that the user requires, the difference in the number of missing values, the columns imputed, the method and range of discretization.</div>
	<div>Output:</div>	<div>Reduction in the time and efforts of the user by providing the automated cleaning. The image cleaner and audio cleaner embedded in this system provides/generates clear image and audio after performing analysis</div>
	<div>Merits:</div>	<div>Avoids manual data removal, saving users' time and making the system more user-friendly.</div>
	<div>Demerits:</div>	<div>Increased resource usage. This system performs a lot of mathematical calculations which is a burden to CPU & GPU</div>
	<div>Future Scope:</div>	<div>This model can be enhanced further to detect and eliminate corrupted data.</div>

<div>Literature Survey</div> <div>TITLE AND AUTHOR</div>	<div>Problem Statement:</div>	<div>The data analysis may take several steps to reach certain conclusions. Simple data can be organized very easily, while the complex data requires proper processing.</div>
<div>The art of Data Analysis</div> <div>Muhammad Ibrahim</div>	<div>Proposed Solution:</div>	<div>Various types of calculations and techniques to analyse complex data. Using parameters like correlation, regression, and graphs like bar charts , histogram, pie charts we can easily analyse the given data</div>
	<div>Output:</div>	<div>The data analysis can provide visualised data information and calculate the different parameters uses various techniques.</div>
	<div>Merits:</div>	<div>Easily visualisable data , can derive solutions and summary easily.</div>
	<div>Demerits:</div>	<div>Highly unorganised data would be hard to analyse.</div>
	<div>Future Scope:</div>	<div>With the increase in various data analysis, the field needs more methods to effectively analyse data in a more efficient way.</div>

<div>Literature Survey</div> <div>TITLE AND AUTHOR</div>	<div>Problem Statement:</div>	<p>The success of AI-based technologies depends crucially on trustful and clean data. Research in data cleaning has provided a variety of approaches to address different data quality problems. Most of them require some prior knowledge about the dataset in order to select and configure the approach correctly. This incremental approach poses the challenge of identifying the right sequence of cleaning routines and their configurations.</p>
<div>Towards Automated Data Cleaning Workflows</div> <div>Mohammad Mahdavi Felix Neutatz Larysa Visengeriyeva Ziawasch Abedjan</div>	<div>Proposed Solution:</div>	<p>Presented their vision and initial steps for supporting the user in building complex pipelines of automated data cleaning tools.</p>
	<div>Output:</div>	<p>Knowledge about pipelines of automated data cleaning tools</p>
	<div>Merits:</div>	<p>Data cleaning made easier.</p>
	<div>Demerits:</div>	<p>Only helps in cleaning data not analysing and visualisation</p>
	<div>Future Scope:</div>	<p>Error detection and correction.</p>

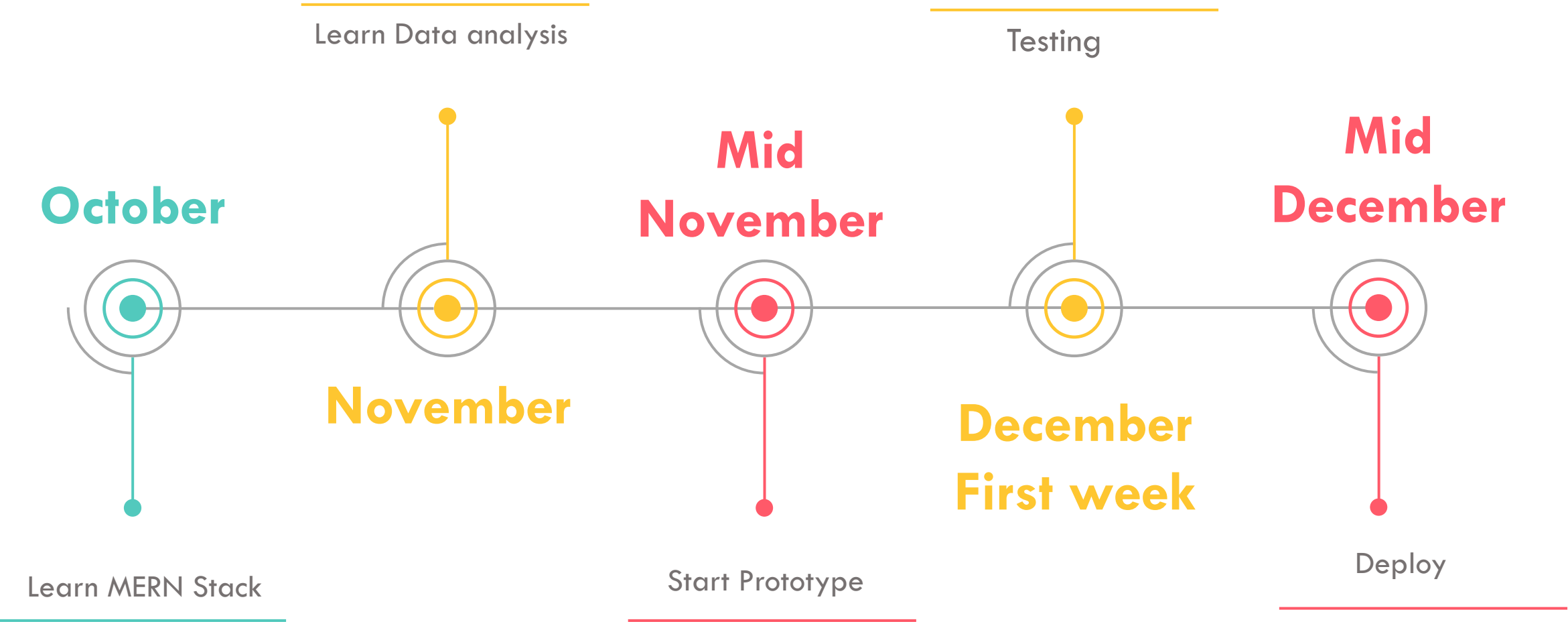
Inference

Microsoft Power BI works by connecting data sources and providing a dashboard of business intelligence to the users. It can connect with just an Excel spreadsheet or bring together cloud-based and on-premises data warehouses. Data pulled from cloud-based sources, such as Salesforce CRM, is automatically refreshed.

Proposed Solution

Having clean data will eventually boost overall productivity and provide for the greatest quality information in your decision-making, which takes time. As a result, an automatic data cleansing tool is necessary. A web-based application that perform analyses on user data on the cloud regardless of the user's computer resources.

Gantt Chart



References

- The art of Data Analysis 1. Muhammad Ibrahim, Department of Statistics, Govt. MAO College, Lahore https://www.researchgate.net/publication/283269432_The_art_of_Data_Analysis
- Characterizing Exploratory Visual Analysis: A Literature Review and Evaluation of Analytic Provenance in Tableau, Leilani Battle¹ and Jeffrey Heer²,
https://homes.cs.washington.edu/~leibatt/static/papers/battle_eurovis2019_cr_04_11_2019.pdf
- Research Data Analysis with Power BI, Vijay Krishnan S Bharanidharan G Krishnamoorthy,
<https://ir.inflibnet.ac.in/bitstream/1944/2116/1/24.pdf>
- Towards Automated Data Cleaning Workflows, Ziawasch Abedjan Felix Neutatz Mohammad Mahdavi Larysa Visengeriyeva ,
https://www.researchgate.net/publication/335136628_Towards_Automated_Data_Cleaning_Workflows
- Automated data cleaning, Jowin Jestine Chiriyankandath,
https://www.researchgate.net/publication/352063621_AUTOMATED_DATA_CLEANING

Thank you 😊!!